Dynamically Consistent Preferences Under Imprecise Probabilistic Information

Frank Riedel[a,1], Jean-Marc Tallon[b,2], Vassili Vergopoulos[c,1]

[a] *Center for Mathematical Economics, Universität Bielefeld Postfach 10 01 31 D-33501 Bielefeld, Germany and University of Johannesburg*
[b] *Paris School of Economics, CNRS*
[c] *Paris School of Economics, University of Paris 1 Panthéon-Sorbonne*

Abstract

This paper extends decision theory under imprecise probabilistic information to dynamic settings. We explore the relationship between the given objective probabilistic information, an agent's subjective multiple priors, and updating. Dynamic consistency implies rectangular sets of priors at the subjective level. As the objective probabilistic information need not be consistent with rectangularity at the subjective level, agents might select priors outside the objective probabilistic information while respecting the support of the given set of priors. Under suitable additional axioms, the subjective set of priors belongs to the rectangular hull of the objective probabilistic information.

Keywords: Ambiguity, Time consistency, Imprecise Probabilistic Information
JEL classification: D81.

1. Introduction

Economic decisions are often made with imprecise knowledge of the statistical properties of the environment, i.e., in a situation of Knightian uncertainty. This Knightian uncertainty need not be absolute, however, as some information about possible probability distributions is usually available. In one of the famous Ellsberg [2] experiments, for example, the agent knows that the probability of drawing a red ball is one-third whereas the probability of drawing a yellow ball is anything between zero and two-thirds. Under such conditions, the agent is faced with a bet (or act) that depends on the outcome of the experiment and some imprecise information about possible probability distributions that can be described by an *information set* (of second order) that contains all objectively possible distributions.

Gajdos, Hayashi, Tallon, and Vergnaud [6] (henceforth GHTV) adapted the basic analysis of Gilboa and Schmeidler [7] (who focus solely on uncertain acts without considering

Email address: `frank.riedel@uni-bielefeld.de` (Frank Riedel)

[1]Corresponding Author; financial support through the German Research Foundation Grant Ri-1128-6-1 is gratefully acknowledged.

[2]Financial support through the Grant ANR-12-FRAL-0008-01 is gratefully acknowledged.

information about possible probability distributions) to such uncertain environments. In this paper, we extend the axiomatic analysis of preferences under Knightian uncertainty with imprecise probabilistic information to dynamic settings.

According to GHTV, an agent who is confronted with an information set of possible priors selects a subjective set of priors and computes the worst expected utility of an act over this set of selected priors. These selected priors are consistent with the given information in the sense that their support is included in the support of objective information.

In dynamic environments, agents need to update their expectations upon the arrival of new information. Epstein and Schneider [4] have shown that it is possible to maintain dynamic consistency for preferences over acts in a multiple-prior setting if agents update their priors in a Bayesian manner prior by prior and if the subjective set of priors is stable under pasting conditional and marginal probabilities from different priors to the original set (or, as Epstein and Schneider call it, rectangular).

In contrast to Epstein and Schneider's setting, an agent is faced with objective yet imprecise information about possible probability distributions in our setting. Ex ante, there is no reason to assume that this information is given by a rectangular set of priors. It is thus not clear how an agent should process such information or whether it is possible to maintain dynamic consistency at all.

We show here that utility functionals in the form of GHTV are dynamically consistent if the subjective set of priors is selected and updated in a suitable way. In the first place, as in GHTV's static analysis, the support of the selected set of priors has to be included in the support of the information set (i.e., the objectively known set of possible priors). In the second place, the initially chosen set has to be stable under pasting, and once the initial set of priors has been chosen, agents update their beliefs prior by prior.

An important element of our analysis is the fact that the subjective set of priors can be larger than the exogenously given set of possible priors because the agent does not want to exclude possible conditional beliefs ex ante yet also wants to be dynamically consistent. The potential "overselection" of priors is an important – albeit necessary – feature of our model.

The overselection should not be too arbitrary, though. In addition to the natural requirement that the selected sets of priors be consistent with the support of objective information, we adapt two further axioms from GHTV to our dynamic setting, namely Reduction (under precise information) and Local Dominance. Local Dominance applies GHTV's dominance concept locally at each node to the next time step. When two acts are resolved in the next period and if one act is preferable to another under every element of the information set, the ranking is unchanged under the whole information set. The reduction axiom states that when the objective information consists of a single prior and this single prior is consistent with the given state of the world in that it puts mass one on the currently observed event, then the agent selects exactly this prior to evaluate acts.

The two axioms of Reduction and Local Dominance force the selected priors to be contained in the rectangular hull of the information set, i.e., the smallest rectangular set containing the initially given probabilistic information. As a first consequence, the overselection of priors does not occur in situations in which the probabilistic information is already rectangular. In other words, the overselection only emerges when the probabilistic information and the filtration are not "well-adapted" to each other. This overslection can thus be seen as an attempt on the agent's part to deal with discrepant sources of information. Second, requiring the subjective set of priors to be in the rectangular hull of the objectively given set of priors implies that Bayesian updates of the initial set of subjective priors belong to updates to the information set. In this sense, no further overselection arises at conditional stages.

Overselection is a sign of the ambiguity-averse decision maker sophistication when one is confronted with a potential conflict among sources of information that would lead to dynamically inconsistent choices. Recognizing the dynamic consistency problem, the decision maker minimally adapts his or her ex-ante preferences so as not to face consistency problems later on yet does not dismiss the ambiguous nature of the situation altogether.

There is usually a tension between dynamic consistency and deviations from expected utility. In fact, Epstein and LeBreton [3] have shown that in order to maintain dynamic consistency along all possible information flows, it is necessary to fall back on a model of probabilistic sophistication that precludes any sensitivity to ambiguity. When the information flow is given, however, it is possible to maintain dynamic consistency for multiple priors and other ambiguity-averse models. Epstein and Schneider [4] (see also Riedel [12]) have shown that multiple prior preferences are dynamically consistent if each prior is updated in a Bayesian way and the set of priors is rectangular or stable under pasting marginal and conditional probabilities. Maccheroni, Marinacci, and Rustichini [10] and Föllmer and Penner [5] have generalized dynamic consistency to variational preferences by characterizing the suitable penalty functions for this large class of ambiguity-averse preferences.

Pires [11] is able to consider updating for arbitrary events by weakening the notion of dynamic consistency. Siniscalchi [14] considers a version of Strotz's (1955) Consistent Planning in which inconsistent agents play a dynamic game against themselves. Hanany and Klibanoff [8] maintained both Dynamic Consistency and Relevance[3] but allowed dynamic preferences and updated sets of priors to depend on the set of feasible plans of actions, some particular optimal plan of actions within this set, and the event that is observed. Hill [9] observed that the incompatibility between Dynamic Consistency and Relevance only holds over "objective trees" and does not preclude their compatibility over "subjective trees". By exploiting this idea, he developed a dynamic extension of the multiple prior model in which the ex-ante set of priors is updated on subjective contingencies in a dynamically consistent and relevant way.

The remainder of the paper is organized as follows: Section 2 briefly reviews the GHTV preferences and representation in a static decision environment. Section 3 describes the dynamic decision environment. It first presents a conditional version of Relevance and Dynamic Consistency and their characterization within dynamic GHTV preferences. It then introduces the axioms of Reduction and Local Dominance and shows how they impose restrictions on the overselection of priors. All proofs are gathered in the Appendix.

[3] What we call Relevance is sometimes referred to as *Consequentialism* in the literature. It requires the preference between two acts to only depend on the outcomes of these acts that remain possible given the information flow.

2. The Framework

2.1. Objects of Choice

Consider two nonempty sets, the outcome space $\mathcal{X}$ and the state space S. The state space S is assumed to be finite. Let $\Delta\mathcal{X}$ denote the set of all lotteries with finite support on $\mathcal{X}$. An act is a function $f : S \to \Delta\mathcal{X}$. Let $\mathcal{F}$ stand for the set of all acts. A constant act with $f(s) = l$ for all $s \in S$ and for some lottery $l \in \Delta\mathcal{X}$ is also denoted by l.

Imprecise probabilistic information is modeled by a nonempty closed and convex set P of probability measures on S. A typical element of P will be denoted by p. Let $\mathcal{P}$ stand for the set of all nonempty, closed, and convex sets of probability measures on S. For $P \in \mathcal{P}$, we let $\operatorname{supp} P$ be the support of P which contains all states $s \in S$ such that there exists $p \in P$ with $p(s) > 0$. For a real–valued function $g : S \to \mathbb{R}$ and a probability measure p on S, we denote by

$$\mathbb{E}_p(g) = \sum_{s \in S} p(s)g(s)$$

the expectation of g under p.

Following GHTV, we consider uncertain acts in conjunction with imprecise probabilistic information as the basic objects of choice. Thus, an agent has a preference relation $\succsim$ defined on $\mathcal{P} \times \mathcal{F}$, the set of pairs of imprecise probabilistic information and acts. For $P, Q \in \mathcal{P}$ and $f, g \in \mathcal{F}$, the preference ranking $(P, f) \succsim (Q, g)$ means that the agent prefers act f under probabilistic information P to act g under probabilistic information Q.

2.2. Static representation

We recall the static representation result of GHTV for a preference relation $\succsim$ on $\mathcal{P} \times \mathcal{F}$. Consider the following list of axioms.

Order $\succsim$ is complete and transitive. As usual, $\succ$ denotes the strict preference relation derived from $\succsim$, and $\sim$ the indifference relation.

Act Continuity For any $P \in \mathcal{P}$ and $f, g, h \in \mathcal{F}$, if $(P, f) \succ (P, g) \succ (P, h)$, there exists $\alpha, \beta \in (0, 1)$ such that $(P, \alpha f + (1-\alpha)h) \succ (P, g) \succ (P, \beta f + (1-\beta)h)$.

Outcome Preference For every $P, Q \in \mathcal{P}$ and $l \in \Delta\mathcal{X}$, $(P, l) \sim (Q, l)$.

Nontriviality There exist $P \in \mathcal{P}$ and $l, m \in \Delta\mathcal{X}$ such that $(P, l) \succ (P, m)$.

C-independence For any $P \in \mathcal{P}$, $f, g \in \mathcal{F}$, $l \in \Delta\mathcal{X}$ and $\lambda \in (0, 1)$, if $(P, f) \succsim (P, g)$, then $(P, \lambda f + (1-\lambda)l) \succsim (P, \lambda g + (1-\lambda)l)$.

Uncertainty Aversion For any $P \in \mathcal{P}$, $f, g \in \mathcal{F}$ and $\lambda \in (0, 1)$, if $(P, f) \sim (P, g)$, then $(P, \lambda f + (1-\lambda)g) \succsim (P, f)$.

Monotonicity For $P \in \mathcal{P}$ and $f, g \in \mathcal{F}$, if $(P, f(s)) \succsim (P, g(s))$ for all $s \in \operatorname{supp} P$, then $(P, f) \succsim (P, g)$.

A binary relation $\succsim$ on $\mathcal{P} \times \mathcal{F}$ is said to be *imprecision averse* or *GHTV* if it satisfies the axioms Order, Act Continuity, Outcome Preference, Nontriviality, C-independence, Uncertainty Aversion and Monotonicity.

The representation theorem states that an imprecision averse decision maker uses a function $\varphi : \mathcal{P} \to \mathcal{P}$ which chooses for an objectively given set of possible distributions P a set of prior $\varphi(P)$ that the agent uses to evaluate the outcomes.

Definition 1. *A mapping $\varphi : \mathcal{P} \to \mathcal{P}$ is support–preserving if $\operatorname{supp} \varphi(P) \subseteq \operatorname{supp} P$ holds for all $P \in \mathcal{P}$.*

Theorem 1 (Gajdos, Hayashi, Tallon, and Vergnaud [6]). *A binary relation $\succsim$ on $\mathcal{P} \times \mathcal{F}$ is* GHTV *if and only if there exist a nonconstant linear utility function $u : \Delta\mathcal{X} \to \mathbb{R}$ and a support–preserving mapping $\varphi : \mathcal{P} \to \mathcal{P}$ such that, for any $P, Q \in \mathcal{P}$ and $f, g \in \mathcal{F}$:*

$$(P, f) \succsim (Q, g) \iff \min_{p \in \varphi(P)} \mathbb{E}_p(u \circ f) \geq \min_{p \in \varphi(Q)} \mathbb{E}_p(u \circ g) \tag{1}$$

In this representation, u is unique up to positive affine transformations, and φ is unique.

Theorem 1 provides a decision-theoretic foundation to the idea that the set of priors is fully determined as a function φ of the objective probabilistic information. In the static setting, the only restriction that φ must satisfy is the support–preserving property; that is, only states that are deemed possible by the probabilistic information can be also deemed possible by the set of priors that the agent selects.

3. Updating and Dynamic Representations

3.1. Time and Information Flow

We consider a discrete time framework with points in time $t = 0, \ldots, T$. The information flow is given by a sequence of partitions $(\pi_t)_{t=0,\ldots,T}$ on S where π_{t+1} refines π_t for any $t = 0, \ldots, T-1$. We assume that $\pi_0 = \{S\}$. For a state $s \in S$ and a time $t = 0, \ldots T$, we denote by $\pi_t(s)$ the unique set in π_t which contains s. We assume that the true state of the world is revealed at time T. Thus, $\pi_T(s) = \{s\}$ for any $s \in S$.

The following example illustrates our dynamic choice framework under imprecise probabilistic information; it will be used throughout the paper.

Example 1. *Consider the following dynamic variant of the Ellsberg [2] experiment. The state space is $S = \{r, b, g\}$ for three possible distinct outcomes "red", "blue", "green". Objective information takes the form of a set of probability measures $P_{a,b} = \{(1/3, p, 2/3 - p) \colon a \leq p \leq b\}$ for some a, b with $0 \leq a \leq b \leq 2/3$. The agent thus knows that the probability for "red" is $1/3$, but she has imprecise information about the odds for "blue" resp. "g". There are three points in time. Ex ante, the agent has no information. At the interim stage, the agent will be told whether the outcome is "green" or not. In the last stage, all information is revealed. Thus, the information flow is determined by the sequence of partitions $\pi_0 = \{S\}$, $\pi_1 = \{E, F\}$, with $E = \{r, b\}$ and $F = \{g\}$, and $\pi_2 = \{\{r\}, \{b\}, \{g\}\}$.*

3.2. Basic Pointwise Representation

We consider a family of preference relations $(\succsim_{t,s})_{t=0,\dots,T,s\in S}$ on $\mathcal{P}\times\mathcal{F}$. For $P,Q\in\mathcal{P}$, $f,g\in\mathcal{F}$, $t=0,\dots,T$ and $s\in S$, the preference ranking $(P,f)\succsim_{t,s}(Q,g)$ means that the agent prefers act f under probabilistic information P to act g under probabilistic information Q conditional upon the event $\pi_t(s)$ that she observes at time t and state s.

As usual for dynamic models of preferences, one can think of each preference relation in the family $(\succsim_{t,s})_{t=0,\dots,T,s\in S}$ in different ways. First, each $\succsim_{t,s}$ can be understood as a revealed preference observed from choices made at time t and information set $\pi_t(s)$. An analyst can indeed elicit the various $\succsim_{t,s}$ as long as the elicitation experiments are repeated at any of the various states s. Second, each $\succsim_{t,s}$ can represent the preference the agent verbally reports to the analyst he expects to have at (t,s). In a more normative perspective, each $\succsim_{t,s}$ represents the preference the agent expects *for himself* to have at (t,s).

There is a natural naive way of specifying conditional preferences $(\succsim_{t,s})_{t=0,\dots,T,s\in S}$ in our framework when an ex ante preference $\succsim$ is given. Consider a set $P\in\mathcal{P}$. Fix $t\in\{0,\dots,T\}$ and $s\in S$. If there exist $p\in P$ such that $p(\pi_t(s))>0$, then we consider the set of Bayesian updates

$$P_t(s)=\{p(\cdot|\pi_t(s)) : p\in P,\ p(\pi_t(s))>0\}\in\mathcal{P}\,.$$

For any $P,Q\in\mathcal{P}$ and $f,g\in\mathcal{F}$, and for any $t=0,\dots,T$ and $s\in S$ such that there exist $p\in P$ and $q\in Q$ with $p(\pi_t(s))>0$ and $q(\pi_t(s))>0$, we define a conditional preference relation as

$$(P,f)\ \succsim_{t,s}\ (Q,g) \quad\Longleftrightarrow\quad (P_t(s),f)\ \succsim\ (Q_t(s),g). \tag{2}$$

We shall see that this naive way of updating generally leads to dynamically inconsistent preferences.

We will use the following axioms.

GHTV For $t\in\{0,\dots,T\}$ and $s\in S$, the binary relation $\succsim_{t,s}$ on $\mathcal{P}\times\mathcal{F}$ is GHTV.

Adaptedness For $t\in\{0,\dots,T\}$ and $s,s'\in S$, if $\pi_t(s)=\pi_t(s')$, then $\succsim_{t,s}\ =\ \succsim_{t,s'}$.

From the axiom GHTV and Theorem 1, we obtain a family $(\varphi_{t,s})_{t=0,\dots,T,s\in S}$ of support–preserving functions $\varphi_{t,s}:\mathcal{P}\to\mathcal{P}$ and a family of nonconstant, linear (Bernoulli) utility functions $(u_{t,s})_{t=0,\dots,T,s\in S}$ such that the preference relation $\succsim_{t,s}$ can be represented by the utility function

$$U_{t,s}(P,f)=\min_{p\in\varphi_{t,s}(P)}\mathbb{E}_p(u_{t,s}\circ f)\,.$$

The axiom Adaptedness requires preferences at time t and state s to only depend upon s through the event $\pi_t(s)$ which the agent observes. Thus, if the agent observes the same event at two different states, the corresponding preferences at that time must be the same. A collection of functions $(\varphi_{t,s})_{t=0,\dots,T,s\in S}$ from $\mathcal{P}$ to $\mathcal{P}$ is said to be *adapted* if, for $t\in\{0,\dots,T\}$ and $s,s'\in S$ such that $\pi_t(s)=\pi_t(s')$, we have $\varphi_{t,s}=\varphi_{t,s'}$.

Thus, for an adapted family, $(\varphi_{t,s})_{t=0,\dots,T,s\in S}$, the priors selected at time t and state s only depend upon the event which the agent observes at (t,s), as well as the available probabilistic information.

We also require that the preferences over sure acts do not change over time and states.

Stable Tastes For $t, t' \in \{0, \ldots, T\}$, $s, s' \in S$, $P, Q \in \mathcal{P}$ and $l, m \in \Delta\mathcal{X}$, we have $(P, l) \succsim_{t,s} (Q, m)$ if and only if $(P, l) \succsim_{t',s'} (Q, m)$.

As a consequence of the axiom "Stable States", the (Bernoulli) utility functions at the various times and states are positive affine transformations of each other. We can thus choose one common (Bernoulli) utility function which we simply denote by u. Moreover, note that the set $\varphi_{0,s}(P)$ of priors selected *ex ante* at any state $s \in S$ is independent of s by Adaptedness and the triviality of π_0. We use the notation $\varphi(P)$ to refer to any of the $\varphi_{0,s}(P)$ for $s \in S$.

3.3. Conditional Relevance

It is natural to require that the agent's preferences do not depend on states which can be excluded at some given point in time. At time t and state s, the agent knows that the event $\pi_t(s)$ happened. Given some probabilistic information, she must then be indifferent between any two acts that agree with each other on $\pi_t(s)$.

For any $P \in \mathcal{P}$ and $A \subseteq S$, we say that A is P–negligible if we have $p(A) = 0$ for all $p \in P$.

Conditional Relevance For $t \in \{0, \ldots, T\}$ and $s \in S$, $P \in \mathcal{P}$, and $f, g \in \mathcal{F}$: if $\pi_t(s)$ is not P–negligible and $f(s') = g(s')$ for all $s' \in \pi_t(s)$, then $(P, f) \sim_{t,s} (P, g)$.

We then obtain that the agent chooses only probability distributions which put full mass on the current information set $\pi_t(s)$ in the sense of the following definition.

Definition 2. *A collection $(\varphi_{t,s})_{t=0,\ldots,T,s\in S}$ of mappings from $\mathcal{P}$ to $\mathcal{P}$ is said to be conditionally relevant if, for $t \in \{0, \ldots, T\}$, $s \in S$, and $P \in \mathcal{P}$ such that $\pi_t(s)$ is not P–negligible, we have $p(\pi_t(s)) = 1$ for all $p \in \varphi_{t,s}(P)$.*

To understand why we restrict Conditional Relevance to situations where the observed event $\pi_t(s)$ is not P-negligible, suppose to the contrary that $\pi_t(s)$ is P-negligible. Then, $\operatorname{supp} P \subseteq S \setminus \pi_t(s)$. By the support-preserving property, we must also have $\operatorname{supp} \varphi_{t,s}(P) \subseteq S \setminus \pi_t(s)$. This implies $p(\pi_t(s)) = 0$ for all $p \in \varphi_{t,s}(P)$ and would contradict the unrestricted versions of both Conditional Relevance and Definition 2. This shows that we need to restrict the agent's preferences only when the two sources of information are not contradictory and justifies that Conditional Relevance has bite only when these two sources are compatible.

To have a more concise language later on, we give the following names to the list of properties for preferences and collections $(\varphi_{t,s})_{t=0,\ldots,T,s\in S}$ of mappings from $\mathcal{P}$ to $\mathcal{P}$.

Definition 3. *We call a family $(\succsim_{t,s})_{t=0,\ldots,T,s\in S}$ of binary relations on $\mathcal{P} \times \mathcal{F}$ that satisfies* GHTV*, Adaptedness, Conditional Relevance, and Stable Tastes a* Conditional Imprecision Averse Preference Family*. We call a family $(\varphi_{t,s})_{t=0,\ldots,T,s\in S}$ of support–preserving, adapted, and conditionally relevant mappings from $\mathcal{P}$ to itself a* prior selection family*.*

Theorem 4 in the Appendix explains in detail the relationship between Conditional Imprecision Averse Preference Families and prior selection families.

3.4. Dynamic Consistency

We now need to connect the priors selected at each state s and time t to one another. We thus introduce Dynamic Consistency to that effect and discuss its consequences in our model.

Dynamic Consistency For $t \in \{0, \ldots, T-1\}$ and $s \in S$, for $P \in \mathcal{P}$ such that $\pi_t(s)$ is not P-negligible, and for $f, g \in \mathcal{F}$: If $(P, f) \succsim_{t+1,s'} (P, g)$ for all $s' \in \pi_t(s)$, then $(P, f) \succsim_{t,s} (P, g)$.

The key insight Dynamic Consistency captures is the following one: at any time $t < T$ and state s, if the possible future preferences at $t+1$ unanimously rank an alternative above another one, then preferences at (t, s) must also rank the former one above. But note that we only require this in situations when $\pi_t(s)$ is not P-negligible; that is, when the realized event and the probabilistic information are coherent with each other.

Dynamic consistency of multiple prior representations has been studied extensively in recent years; after the basic insight of Sarin and Wakker [13] of the role of rectangularity in a two period example, Epstein and Schneider [4] characterize dynamic consistency for intertemporal consumption choice problems, and Delbaen [1] and Riedel [12] achieve the same for dynamic risk measures.

Consider a probability measure p on S. Fix $t \in \{0, \ldots, T\}$ and $s \in S$. If $p(\pi_t(s)) > 0$, then, define $p_t(s) = p(\cdot\, |\pi_t(s))$, which is another probability measure. We can also view $p_t(\cdot)$ as a transition kernel on S. Moreover, for $t < T$, define $p_t^{+1}(s)$ as the restriction of $p_t(s)$ to the algebra generated by π_{t+1}.

Consider a set $P \in \mathcal{P}$. Fix $t \in \{0, \ldots, T\}$ and $s \in S$. If $\pi_t(s)$ is not P-negligible, then we define:

$$P_t(s) = \{p_t(s) : p \in P,\ p(\pi_t(s)) > 0\} \in \mathcal{P} \quad \text{and} \quad P_t^{+1}(s) = \{p_t^{+1}(s) : p \in P_t(s)\}.$$

Definition 4. *Fix $t \in \{0, \ldots, T-1\}$ and $s \in S$. For probability measures p, q on S, define the pasting $p \circ_{t,s} q$ of p and q after (t, s) as follows. If $q(\pi_{t+1}(s)) = 0$, set $p \circ_{t,s} q = p$. Otherwise, we set for $s' \in S$*

$$p \circ_{t,s} q(s') = \begin{cases} q(s'|\pi_{t+1}(s))\ p(\pi_{t+1}(s)) & \text{if } s' \in \pi_{t+1}(s) \\ p(s') & \text{else.} \end{cases}$$

For $P, Q \in \mathcal{P}$, we define their pasting after (t, s) to be

$$P \circ_{t,s} Q = \{p \circ_{t,s} q : p \in P, q \in Q\}\ .$$

We call a family $(P_{t,s})_{t=0,\ldots,T, s\in S}$ of sets of priors $P_{t,s} \in \mathcal{P}$ *stable under pasting* (or rectangular) if for all $t = 0, \ldots, T-1$ and $s, s' \in S$ such that $s' \in \pi_t(s)$ we have

$$P_{t,s} \circ_{t,s'} P_{t+1,s'} = P_{t,s}\ .$$

Similarly, we call a prior selection family $(\varphi_{t,s})_{t=0,\ldots,T,s\in S}$ stable under pasting if for all $t = 0, \ldots, T-1$, all $s \in S$ such that $\pi_t(s)$ is not P-negligible and all $s' \in \pi_t(s)$

$$\varphi_{t,s}(P) \circ_{t,s'} \varphi_{t+1,s'}(P) = \varphi_{t,s}(P)\ .$$

The pasting $p \circ_{t,s} q$ of p and q after (t,s) describes a probability distribution whose Bayesian update on $\pi_{t+1}(s)$ agrees with that of q. But its Bayesian update on $S \setminus \pi_{t+1}(s)$, as well as its one-step-ahead restriction to $\{\pi_{t+1}(s), S \setminus \pi_{t+1}(s)\}$, agree with those of p. Such pasting can be extended to sets of probability measures. It is always possible to close a given family $(P_{t,s})_{t=0,\ldots,T,s\in S}$ under pasting; we call the resulting family of priors the rectangular hull of $(P_{t,s})_{t=0,\ldots,T,s\in S}$ and denote it by $(\mathrm{rect}_{t,s}(P))_{t=0,\ldots,T,s\in S}$. More precisely, for any $s \in S$, let δ_s be the degenerate measure assigning a probability of 1 to state s. We then define $\mathrm{rect}_{t,s}(P)$ recursively for all $t = 0, \ldots, T$, $s \in S$ and $P \in \mathcal{P}$ such that $\pi_t(s)$ is not P-negligible by setting

$$\mathtt{rect}_{T,s}(P) \;:=\; P_T(s) \;=\; \{\delta_s\}, \text{ and}$$

$$\mathtt{rect}_{t,s}(P) \;:=\; \left\{ \int_S p(s') \cdot dm(s') : m \in P_t(s)^{+1},\ p(s') \in \mathtt{rect}_{t+1,s'}(P) \right\}.$$

Theorem 2. 1. *A Conditional Imprecision Averse Preference Family* $(\succsim_{t,s})_{t=0,\ldots,T,s\in S}$ *satisfies Dynamic Consistency if and only if there exist a nonconstant linear utility function* $u : \Delta\mathcal{X} \to \mathbb{R}$ *and a prior selection family* $(\varphi_{t,s})_{t=0,\ldots,T,s\in S}$ *that is stable under pasting such that* $\succsim_{t,s}$ *is represented by the utility function*

$$U_{t,s}(P,f) \;=\; \min_{p\in\varphi_{t,s}(P)} \mathbb{E}_p(u \circ f)\,. \tag{3}$$

u is unique up to positive affine transformations; $(\varphi_{t,s})_{t=0,\ldots,T,s\in S}$ *is unique.*

2. *When the equivalent conditions of part 1. hold true, the following additional properties are satisfied.*
   (a) *(Dynamic Programming) For any* $t \in \{0,\ldots,T-1\}$, $s \in S$, $P \in \mathcal{P}$ *such that* $\pi_t(s)$ *is not P-negligible, and any* $f \in \mathcal{F}$,

$$U_{t,s}(P,f) = \min_{m\in\varphi_{t,s}(P)^{+1}} \{ \int_S U_{t+1,s'}(P,f) \cdot dm(s') \}. \tag{4}$$

   (b) *(Full Bayesian Updating) For any* $t \in \{0,\ldots,T-1\}$, $s \in S$ *and* $P \in \mathcal{P}$ *such that* $\pi_t(s)$ *neither P-negligible nor $\varphi(P)$-negligible,*

$$\varphi_{t,s}(P) = \varphi(P) \mid \pi_t(s) \equiv \{p(.|\pi_t(s)) : p(\pi_t(s)) > 0 \text{ and } p \in \mathcal{P}\}\,. \tag{5}$$

A dynamically consistent agent selects priors at time t and state s according to $\varphi_{t,s}$. The priors selected at (t,s) only deem as possible states that are already deemed possible by the probabilistic information itself. Moreover, they only depend on the probabilistic information and the event which she observes at (t,s).

Theorem 2 characterizes Dynamic Consistency for Conditional Imprecision Averse Preference families in terms of the stability under pasting of the prior selection family. Equation (4) shows that Dynamic Consistency leads to value functions $\{U_{t,s},\ t = 0,\ldots,T,\ s \in S\}$ with a recursive structure. This property lies at the heart of dynamic programming methods as it ensures the equality between backward induction solutions and *ex-ante* optimal plans.

Furthermore,Theorem 2 establishes the Full Bayes updating rule as a consequence of Dynamic Consistency: given some objective information P, the priors that are selected at (t,s) consist of the Bayesian updates on the available event $\pi_t(s)$ of all the priors selected *ex ante* under P. Thus, it is still true, as in Theorem 4, that the priors selected at (t,s) only depend on the probabilistic information P and the available event $\pi_t(s)$. Equation (5) clarifies that these priors only depend on P through the set $\varphi(P)$ of priors selected *ex ante*. In this sense, the priors selected at (t,s) can also be seen as fully determined by the *ex-ante* priors (under the same objective information) and the available event.

Theorem 2 leaves a lot of freedom for the choice of $\varphi(P)$. In light of GHTV's Theorem 2, one might expect the *ex-ante* prior selection process to satisfy the following property: $\varphi(P) \subseteq P$ for any $P \in \mathcal{P}$. GHTV call this the selection property. It would mean that the agent selects her *ex-ante* priors within the available objective information. But the next example shows that this is too restrictive as it sometimes implies neutrality to ambiguity, an undesirable feature.

Example 2. *Using the notations from Example 1, fix $a_E, b_E, \underline{m}, \overline{m} \in [0,1]$ possibly depending on a and b with $a_E \leq b_E$ and $\underline{m} \leq \overline{m}$. Define*

$$\varphi_E(P_{a,b}) = \{(p, 1-p, 0),\ a_E \leq p \leq b_E\} \quad \textit{and} \quad \varphi_F(P_{a,b}) = \{(0,0,1)\},$$

$$\varphi_0(P_{a,b}) = \{\left(\left(\frac{1}{3}+m\right)p, \left(\frac{1}{3}+m\right)(1-p), \frac{2}{3}-m\right) \ : a_E \leq p \leq b_E,\ \underline{m} \leq m \leq \overline{m}\}.$$

The collection $\{\varphi_0(P_{a,b}), \varphi_E(P_{a,b})), \varphi_F(P_{a,b}))\}$ is stable under pasting for any a, b, consistently with Theorem 2[4]*.*

It might be tempting to choose $a_E, b_E, \underline{m}, \overline{m} \in [0,1]$ so as to have $\varphi_0(P_{a,b}) \subseteq P_{a,b}$ for any a, b. However, for any mappings φ_0, φ_E and φ_F such that $\{\varphi_0(P_{a,b}), \varphi_E(P_{a,b})), \varphi_F(P_{a,b}))\}$ is stable under pasting and $\varphi_0(P_{a,b}) \subseteq P_{a,b}$, we have that $\varphi_0(P_{a,b})$ is a singleton. Indeed, suppose that $(1/3, p, 2/3-p)$ and $(1/3, q, 2/3-q)$ belong to $\varphi_0(P_{a,b})$. Then, by stability under pasting,

$$\left(\frac{\frac{1}{3}+p}{1+3q}, \frac{3q(\frac{1}{3}+p)}{1+3q}, 2/3-p\right) \;=\; (1/3, p, 2/3-p) \circ_E (1/3, q, 2/3-q) \in \varphi_0(P_{a,b}) \subseteq P_{a,b}.$$

But then $(1/3+p)/(1+3q) = 1/3$ and, therefore, $p = q$. Hence, imposing that the selected priors be a subset of the set P has the overly strong implication in this example to impose that the agent is neutral to the ambiguity of the situation captured by the fact that P is not a singleton.

In general, sets of priors P contain a rectangular subset if and only if they have a nonempty interior in the appropriate parametrization given by marginal and conditional

---

[4]We abuse slightly notation here. Note that time is here always $t = 0$ or $t = 1$. By Stable Tastes, $\varphi_{0,s}$ is independent of s, and denoted $\varphi(P)$. By Adaptedness, $\varphi_{1,s}$ is constant over E (resp. F), and denoted φ_E (resp. φ_F).

probabilities. We illustrate this fact within our example. Every probability measure p on S can be represented by a pair $(c,d) \in [0,1]^2$, where $c = p(\{b,g\})$ and $d = p(\{b\}|\{b,g\})$. Likewise, any set $P \in \mathcal{P}$ can be represented by a subset $A(P) \subseteq [0,1]^2$. It is easy to see that a subset $P \in \mathcal{P}$ is stable under pasting if and only if $A(P) = [\underline{c}, \overline{c}] \times [\underline{d}, \overline{d}]$, for some $\underline{c}, \overline{c}, \underline{d}, \overline{d} \in [0,1]$.

Suppose now that objective information is given by a set $P \in \mathcal{P}$ such that $A(P)$ has empty interior in $[0,1]^2$. Then, if $Q \in \mathcal{P}$ is stable under pasting and included in P, $A(Q)$ must necessarily have an empty interior and thus be of the form $\{c\} \times [\underline{d}, \overline{d}]$ or $[\underline{c}, \overline{c}] \times \{d\}$. This does not force a full neutrality to ambiguity as in the Ellsberg example above, but nonetheless requires one of the likelihood of $\{b,g\}$ or the likelihood of $\{b\}$ conditional on $\{b,g\}$ to be unambiguous. Thus, imposing Dynamic Consistency while requiring selected priors to be contained in objective information proves again to be too restrictive.

Ambiguity is typically attributed to the "poor" quality of the probabilistic information that an agent has. Thus, one could interpret the fact that the selection property implies ambiguity neutrality in cases where the objective information P does not contain nontrivial rectangular subsets as meaning that these sets P represent information of "good" quality, or are "falsely ambiguous". This interpretation however is not satisfactory, as such sets P can still be very dispersed. In the example, $P_{0,\frac{2}{3}}$ contains no nontrivial rectangular subset but is still imprecise enough and generates ambiguity as demonstrated by the Ellsberg [2] paradox itself.

In light of this discussion, one could think of the following procedure: when faced with a set P that is not rectangular, the decision maker first "rectangularizes" the set, and then selects from it. This would impose the condition $\varphi(P) = \varphi(\texttt{rect}(P))$ for any $P \in \mathcal{P}$; the inclusion $\varphi(P) \subseteq P$ would be required only for sets $P \in \mathcal{P}$ that are already rectangular. In this way, the agent always "rectangularizes" the given objective information in order to adapt it to the structure of the information flow. In particular, this procedure only implies that $\varphi(P)$ is always contained in $\texttt{rect}(P)$ and thus allows for nontrivial rectangular sets $\varphi(P)$ of selected priors.

However this procedure has the non-desirable feature that it does not allow to capture some (imprecision averse) behavior. To illustrate this, fix a single partition π. Fix $P, Q \in \mathcal{P}$. In the spirit of GHTV's Definition 2, we say P is conditionally more precise than Q with respect to π if the following conditions hold:

(i) $P \subseteq Q$,

(ii) For all $p \in P$, $q \in Q$ and $E \in \pi$, $p(E) = q(E) > 0$,

(iii) For all $E \in \pi$, $\{p(\cdot|E) : p \in P\} = \{q(\cdot|E) : q \in Q\}$.

If P and Q satisfy these conditions, they have the same rectangular hull, that is, $\texttt{rect}(P) = \texttt{rect}(Q)$. We would then obtain $\varphi(P) = \varphi(Q)$. In turn, given the representation obtained in Theorem 2, this implies that (P, f) and (Q, f) must be indifferent for any f. Thus, imposing that $\varphi(P) = \varphi(\texttt{rect}(P))$ for any $P \in \mathcal{P}$ implies that the decision maker is indifferent to imprecision (in the sense of GHTV's Axiom 10) within the filtration defined by π.

Our view is therefore that overselection of priors, when the original set does not contain nontrivial rectangular subsets, is a desirable feature, as it does not force the decision criterion to be expected utility. Yet the rectangularization procedure is not a satisfactory solution, as it forces a form of indifference to imprecision. We further investigate the way the selection operates in the next section by weakening this procedure.

3.5. Local Dominance

A consequence of Theorem 2 is that an agent with preferences satisfying Dynamic Consistency and Conditional Relevance and revealing a nonneutral attitude towards ambiguity must sometimes select her *ex-ante* priors outside the probabilistic information that she disposes of. So far, the only restriction on this overselection is the support-preserving property of the prior selection family: the priors selected *ex ante* must only assign positive probability weight to states already receiving a positive weight from the probabilistic information. Any prior is thus not allowed. We now develop other restrictions for this overselection.

The Dominance criterion employed by GHTV restricts the agent's choice of priors to subsets of P. We will now impose a weaker and local version of this dominance criterion that is suited to our dynamic framework.

Local Dominance For $t \in \{0, \dots, T-1\}$, $s \in S$, $P \in \mathcal{P}$ such that $\pi_t(s)$ is not P–negligible,and $f, g \in \mathcal{F}$ that are π_{t+1}–measurable: if $(\{p\}, f) \succsim_{t,s} (\{p\}, g)$ for all $p \in P_t(s)$, then $(P, f) \succsim_{t,s} (P, g)$.

Because of the role played by the Bayesian updating of the objective information on the events in the filtration, Local Dominance can be understood as a requirement of consistency between preferences and the two sources of information, the objective probabilistic set and the filtration. Moreover, Local Dominance can be also seen as a criterion of internal consistency of the preference relation $\succsim_{t,s}$ at some pair (t, s): if act f is at least as good as act g under the Bayesian update on $\pi_t(s)$ of any of the probability distributions in P, then f must also be at least as good as g under P. But we only require this for acts f and g are measurable with respect to the partition of the next stage. Omitting this restriction would lead to the inclusion $\varphi(P) \subseteq P$ for any $P \in \mathcal{P}$ (See GHTV's Theorem 2), which as explained would be too strong for our purposes. In fact, the dominance reasoning captured in the axiom becomes questionable when applied to nonmeasurable acts: since the uncertainty attached to these acts is not fully resolved at the next stage, the ambiguities perceived at the disjoint cells of the next stage partition might hedge one another and explain failures of the dominance reasoning. The measurability restriction is thus meant to allow such hedging to play a role in decisions.

We will also use a version of the criterion of Reduction under Precise Information employed by GHTV. But it requires additional notation. Fix $p \in \Delta S$ and $f \in \mathcal{F}$. Then, there is a partition $(E_1, ..., E_n)$ and a collection $(l_1, ..., l_n)$ of lotteries on $\mathcal{X}$ such that $f(s) = l_i$ for any $s \in E_i$ and any $i \in [1, n]$. Then, define $l(p, f) \in \Delta\mathcal{X}$ as the lottery given by $\sum_{i=1}^{n} p(E_i) \cdot l_i$. Note that this definition is independent of the specific partition $(E_1, ..., E_n)$ that is chosen to construct $l(p, f)$.

Reduction For any $t \in \{0 \ldots T-1\}$, any $s \in S$, and for any $p \in \Delta S$ such that $p(\pi_t(s)) = 1$, we have $(\{p\}, f) \sim_{t,s} (\{p\}, l(p, f))$, for any $f \in \mathcal{F}$.

Under the Reduction axiom, whenever the objective information consists of a single probability measure that is consistent with the available event, the selected priors must be that measure as captured by the following definition.

Definition 5. *A prior selection family $(\varphi_{t,s})_{t=0,\ldots,T,s\in S}$ is said to be grounded if $\varphi_{t,s}(\{p\}) = \{p\}$ for any $p \in \Delta S$ such that $p(\pi_t(s)) = 1$, and any $t \in \{0, \ldots, T\}$ and $s \in S$.*

The next theorem uses Local Dominance and Reduction to further constrain the selection of priors.

Theorem 3. *A Conditional Imprecision Averse Preference Family $(\succsim_{t,s})_{t=0,\ldots,T,s\in S}$ satisfies Dynamic Consistency, Local Dominance and Reduction if and only if the prior selection family $(\varphi_{t,s})_{t=0,\ldots,T,s\in S}$ in the representation (3) is stable under pasting, grounded and we have*

$$\varphi_{t,s}(P) \subseteq \mathtt{rect}_{t,s}(P), \tag{6}$$

for all $t \in \{0, \ldots, T\}$, $s \in S$, and $P \in \mathcal{P}$ such that $\pi_t(s)$ is not P–negligible.

In the dynamic setting, the probabilistic information P might be given in a way that does not fit well with the structure of the information flow of states of nature in the sense that P itself is not stable under pasting (or rectangular) with respect to the filtration of events. Equation (6) shows that the agent "rectangularizes" the probabilistic information (or closes it under pasting according to the given information flow). In other words, she chooses freely her priors within the rectangular hull of P. Note that this is always consistent with the support-preserving property of the prior selection family as the rectangularization itself preserves the support of a set of measures.

Confining the additional priors within the rectangular hull of probabilistic information has at least two advantages. First, if the objective information P fits well the structure of the information flow, then P is already rectangular. In particular, we obtain $\varphi(P) \subseteq P$. Thus, the agent is only allowed to select *ex-ante* priors outside the objective information in situations where the latter is not well-adapted to the filtration. Second, the rectangular hull of a set P does never add new posteriors conditional on the events in the filtration. Therefore, at any (t, s) such that $\pi_t(s)$ is a proper and not P–negligible subset of S, we have $\varphi_{t,s}(P) \subseteq P_t(s)$. Thus, all the selected priors must be Bayesian updates on the available event of measures in the objective information. Hence, overselection is here the natural consequence of the decision maker's desire to act in a dynamically consistent manner while acknowledging the imprecision of the information he has.

Let us come back to our example.

Example 3. *With the notations from Example 1, the rectangular hull of the sets $P_{a,b}$ is given by*

$$\boldsymbol{rect}_E(P_{a,b}) = \left\{ \left(\frac{1}{(1+3p)}, \frac{3p}{(1+3p)}, 0 \right) \ : a \leq p \leq b \right\} \quad \text{and} \quad \boldsymbol{rect}_F(P_{a,b}) = \{(0,0,1)\},$$

$$\boldsymbol{rect}_0(P_{a,b}) = \left\{ \left(\frac{\frac{1}{3}+m}{(1+3p)}, \frac{\left(\frac{1}{3}+m\right)3p}{(1+3p)}, \frac{2}{3} - m \right) \ : a \leq m, p \leq b \right\}.$$

Now, for any a, b, let $a', b' \in [0,1]$, possibly depending on a and b, be such that $a \leq a' \leq b' \leq b$. Then, define $(\varphi_0(P_{a,b}), \varphi_E(P_{a,b})), \varphi_F(P_{a,b})))$ according to

$$\varphi_E(P_{a,b}) = \left\{ \left(\frac{1}{(1+3p)}, \frac{3p}{(1+3p)}, 0 \right) \ : a' \leq p \leq b' \right\} \quad \text{and} \quad \varphi_F(P_{a,b}) = \{(0,0,1)\},$$

$$\varphi_0(P_{a,b}) = \left\{ \left(\frac{(\frac{1}{3}+m)}{(1+3p)}, \frac{(\frac{1}{3}+m)3p}{(1+3p)}, \frac{2}{3} - m \right) \ : a' \leq m, p \leq b' \right\}.$$

By construction, $\{\varphi_0(P_{a,b}), \varphi_E(P_{a,b})), \varphi_F(P_{a,b}))\}$ is stable under pasting and satisfies Equation (6) for any a, b, consistently with Theorem 3.

Appendix

Proof of Theorem 2

The next theorem summarizes the discussion from Sections 3.2 and 3.3 and is given without a proof.

Theorem 4. *A collection $(\succsim_{t,s})_{t=0,\dots,T,s\in S}$ is a Conditional Imprecision Averse Preference Family if and only if there exist a nonconstant linear utility function $u : \Delta\mathcal{X} \to \mathbb{R}$ and a prior selection family $(\varphi_{t,s})_{t=0,\dots,T,s\in S}$ such that $\succsim_{t,s}$ is represented by the utility function*

$$U_{t,s}(P,f) \ = \min_{p\in\varphi_{t,s}(P)} \mathbb{E}_p(u \circ f)\,. \tag{7}$$

Moreover, u is unique up to positive affine transformations, and $(\varphi_{t,s})_{t=0,\dots,T,s\in S}$ is unique.

We now turn at the proof of Theorem 2. By Theorem 4, there exist a nonconstant linear utility function $u : \Delta\mathcal{X} \to \mathbb{R}$ and a family $(\varphi_{t,s})_{t=0,\dots,T,s\in S}$ of support–preserving, adapted, and conditionally relevant mappings from $\mathcal{P}$ to itself such that for any $t \in \{0,\dots,T\}$ and $s \in S$, and for any $P, Q \in \mathcal{P}$ and $f, g \in \mathcal{F}$:

$$(P,f) \succsim_{t,s} (Q,g) \Longleftrightarrow U_{t,s}(P,f) \geq U_{t,s}(Q,g), \tag{8}$$

where, for any $P \in \mathcal{P}$ and $f \in \mathcal{F}$, we have:

$$U_{t,s}(P,f) = \min_{p\in\varphi_{t,s}(P)} \mathbb{E}_p(u \circ f) \tag{9}$$

Now, fix $t \in \{0,\ldots,T-1\}$, $s \in S$ and $P \in \mathcal{P}$ such that $\pi_t(s)$ is not P-negligible. Moreover, fix $f \in \mathcal{F}$. Note that it is simple to obtain the following fact: For $t' \in \{0,\ldots,T\}$ and $s' \in S$, there exists $l_{t',s'} \in \Delta\mathcal{X}$ such that $(P,f) \sim_{t',s'} (P, l_{t',s'})$. We can further assume without loss of generality that we have $l_{t',s'} = l_{t'',s''}$ whenever $\succsim_{t',s'} = \succsim_{t'',s''}$. Let us consider the act $g \in \mathcal{F}$ defined by $g(s') = l_{t+1,s'}$ for any $s' \in S$. Since $\pi_t(s)$ is not P-negligible, it is also true that $\pi_t(s')$ is not P-negligible for any $s' \in \pi_t(s)$. We can apply Conditional Relevance and obtain $(P,g) \sim_{t+1,s'} (P, l_{t+1,s'})$ for any $s' \in \pi_t(s)$. This gives $(P,g) \sim_{t+1,s'} (P,f)$ for any $s' \in \pi_t(s)$. Then, by Dynamic Consistency, $(P,g) \sim_{t,s} (P,f)$. Therefore, by Equation (8), and using the π_{t+1}-measurability of g,

$$\mathtt{min}_{p\in\varphi_{t,s}(P)}\mathbb{E}_p(u\circ f) = \mathtt{min}_{p\in\varphi_{t,s}(P)}\mathbb{E}_p(u\circ g) = \mathtt{min}_{m\in\varphi_{t,s}(P)^{+1}}\{\int_S u(l_{t+1,s'})\cdot dm(s')\}. \quad (10)$$

Meanwhile, given the definition of $l_{t+1,s'}$ as well as the representation obtained in Equation (8), we have

$$u(l_{t+1,s'}) = \mathtt{min}_{p\in\varphi_{t+1,s'}(P)}\mathbb{E}_p(u\circ f) \quad (11)$$

Therefore, combining Equation (10) and (11), we obtain:

$$\mathtt{min}_{p\in\varphi_{t,s}(P)}\mathbb{E}_p(u\circ f) = \mathtt{min}_{m\in\varphi_{t,s}(P)^{+1}}\{\int_S \mathtt{min}_{p\in\varphi_{t+1,s'}(P)}\mathbb{E}_p(u\circ f)\cdot dm(s')\} \quad (12)$$

Hence, we have the dynamic programming principle of Equation (4). As a consequence, we can now also write

$$\mathtt{min}_{p\in\varphi_{t,s}(P)}\mathbb{E}_p(u\circ f) = \mathtt{min}_{p\in\tilde{\varphi}_{t,s}(P)}\mathbb{E}_p(u\circ f), \quad (13)$$

where $\tilde{\varphi}_{t,s}(P)$ is the closed and convex set defined by

$$\tilde{\varphi}_{t,s}(P) = \{\int_S p(s')\cdot dm(s') : m \in \varphi_{t,s}(P)^{+1},\ p(s') \in \varphi_{t+1,s'}(P)\ \}. \quad (14)$$

By the uniqueness part of the Gilboa and Schmeidler [7] theorem, we obtain $\varphi_{t,s}(P) = \tilde{\varphi}_{t,s}(P)$. As a result, for any $t \in \{0,\ldots,T-1\}$, $s \in S$ and $P \in \mathcal{P}$ such that $\pi_t(s)$ is not P-negligible, we have

$$\varphi_{t,s}(P) = \{\int_S p(s')\cdot dm(s') : m \in \varphi_{t,s}(P)^{+1},\ p(s') \in \varphi_{t+1,s'}(P)\ \}. \quad (15)$$

Furthermore, for any state $s' \in \pi_t(s)$ such that $\pi_{t+1}(s')$ is not $\varphi_{t,s}(P)$-negligible, Equation (15) gives

$$\varphi_{t+1,s'}(P) = \varphi_{t,s}(P) \mid \pi_{t+1}(s'). \quad (16)$$

We now show that the prior selection family $(\varphi_{t,s})_{t=0,\ldots,T,s\in S}$ is stable under pasting. Let $t \in \{0,\ldots,T-1\}$, $s \in S$ and $P \in \mathcal{P}$ such that $\pi_t(s)$ is not P-negligible. It is sufficient to show $\varphi_{t,s}(P) \circ_{t,s'} \varphi_{t+1,s'}(P) = \varphi_{t,s}(P)$ for any $s' \in \pi_t(s)$.

First, take $p \in \varphi_{t,s}(P)$. If $p(\pi_{t+1}(s')) > 0$, then $\pi_{t+1}(s')$ is not $\varphi_{t,s}(P)$-negligible and, by Equation (16), we have $p(.|\pi_{t+1}(s')) \in \varphi_{t+1,s'}(P)$ and $p = p \circ_{t,s'} p(.|\pi_{t+1}(s')) \in \varphi_{t,s}(P) \circ_{t,s'}$

$\varphi_{t+1,s'}(P)$. If $p(\pi_{t+1}(s'))=0$. Then, $p=p\circ_{t,s'}q$ for any measure q on S. So it is sufficient to take $q\in\varphi_{t+1,s'}(P)$ to obtain $p\in\varphi_{t,s}(P)\circ_{t,s'}\varphi_{t+1,s'}(P)$.

Now, take $p\in\varphi_{t,s}(P)\circ_{t,s'}\varphi_{t+1,s'}(P)$. So $p=m\circ_{t,s'}q$ with $m\in\varphi_{t,s}(P)$ and $q\in\varphi_{t+1,s'}(P)$. Then, p must be an element of $\tilde{\varphi}_{t,s}(P)$ and, therefore, of $\varphi_{t,s}(P)$.
Moreover, an induction on $t\in\{0,\ldots,T\}$ relying upon Equation (16) finally shows the following equality: for any $t\in\{0,\ldots,T\}$, $s\in S$ and $P\in\mathcal{P}$ such that $\pi_t(s)$ neither P-negligible nor $\varphi(P)$-negligible,

$$\varphi_{t,s}(P)=\varphi(P)\mid\pi_t(s), \tag{17}$$

Now assume that the prior selection family $(\varphi_{t,s})_{t=0,\ldots,T,s\in S}$ in the representation (3) is stable under pasting. Fix $t\in\{0,\ldots,T-1\}$, $s\in S$ and $P\in\mathcal{P}$ such that $\pi_t(s)$ is not P-negligible. By iterative applications of stability under pasting, we obtain the equality $\varphi_{t,s}(P)=\tilde{\varphi}_{t,s}(P)$. This entails in turn that Equation (12) holds for any $f\in\mathcal{F}$. From there, dynamic consistency easily follows.

Proof of Theorem 3

First assume Dynamic Consistency, Local Dominance and Reduction. Let u, $(\varphi_{t,s})_{t=0,\ldots,T,s\in S}$ and $(U_{t,s})_{t=0,\ldots,T,s\in S}$ be as in Theorem 2. Let us now use Reduction to show groundedness; that is, we show that $\varphi_{t,s}(\{p\})=\{p\}$ for any $p\in\Delta S$ such that $p(\pi_t(s))=1$, and any $t\in\{0,\ldots,T\}$ and $s\in S$. By Reduction and linearity, for any $f\in\mathcal{F}$, $U_{t,s}(\{p\},f)=U_{t,s}(\{p\},l(p,f))=u(l(p,f))=\sum_{i=1}^{n}u(l_i)p(E_i)=\mathbb{E}_p(u\circ f)$. Therefore, by the uniqueness part of Gilboa and Schmeidler's theorem (1989), we obtain $\varphi_{t,s}(\{p\})=\{p\}$.
Now, fix $P\in\mathcal{P}$ and $s\in S$ such that $P_T(s)\neq\varnothing$. We show that $\varphi_{T,s}(P)\subseteq P_T(s)$. Since $P_T(s)\neq\varnothing$, we have that $\pi_T(s)=\{s\}$ is not P-negligible. Since the prior selection family is conditionally relevant, any prior in $\varphi_{T,s}(P)$ puts a probability of 1 on $\{s\}$. Therefore, $\varphi_{T,s}(P)$ only contains the Dirac distribution at s. Similarly for $P_T(s)$. Hence the inclusion $\varphi_{T,s}(P)\subseteq P_T(s)$.
Moreover, we show that, for any $P\in\mathcal{P}$, $s\in S$ and $t<T$ such that $P_t(s)\neq\varnothing$, we have $\varphi_{t,s}(P)^{+1}\subseteq P_t^{+1}(s)$ by means of contradiction. If $\varphi_{t,s}(P)^{+1}\not\subseteq P_t^{+1}(s)$, then there exists $p^*\in\varphi_{t,s}(P)^{+1}$ such that $p^*\notin P_t^{+1}(s)$. By the separation theorem, we obtain an π_{t+1}-measurable function $F:S\to\mathbb{R}$ such that:

$$\mathtt{min}_{p\in\varphi_{t,s}(P)^{+1}}\mathbb{E}_p(F)\leq\mathbb{E}_{p^*}(F)<\mathtt{min}_{p\in P_t^{+1}(s)}\mathbb{E}_p(F) \tag{18}$$

Without loss of generality, we assume that F is of norm less than 1. By adequately normalizing u if necessary, we can also assume that the range of u contains $[-1,1]$. So F is necessarily of the form $F=u\circ f$, for some π_{t+1}-measurable $f\in\mathcal{F}$. Therefore, Equation (18) becomes:

$$U_{t,s}(P,f)\leq\mathbb{E}_{p^*}(u\circ f)<\mathtt{min}_{p\in P_t^{+1}(s)}\mathbb{E}_p(u\circ f) \tag{19}$$

Now, define another π_{t+1}-measurable $g=(1/2)f+(1/2)l\in\mathcal{F}$ where $l\in\Delta\mathcal{X}$ is defined by $l=l(p^*,f)$ and, therefore, satisfies:

$$u(l)=\mathbb{E}_{p^*}(u\circ f) \tag{20}$$

On the one hand, for any $p \in P_t(s)$, we have $\varphi_{t,s}(\{p\}) = \{p\}$. So $U_{t,s}(\{p\}, g) = \mathbb{E}_p(u \circ g) = (1/2)\mathbb{E}_p(u \circ f) + (1/2)u(l) = (1/2)\mathbb{E}_p(u \circ f) + (1/2)\mathbb{E}_{p^*}(u \circ f) \leq \mathbb{E}_p(u \circ f) = U_{t,s}(\{p\}, f)$ by Equation (19). Thus, $(\{p\}, f) \succsim_{t,s} (\{p\}, g)$ for any $p \in P_t(s)$.
On the other hand, $U_{t,s}(P, g) = (1/2)U_{t,s}(P, f) + (1/2)u(l) = (1/2)U_{t,s}(P, f) + (1/2)\mathbb{E}_{p^*}(u \circ f) \geq U_{t,s}(P, f)$ by Equation (19). Thus, $(P, g) \succsim_{t,s} (P, f)$. But then Local Dominance is contradicted, which finally shows that $\varphi_{t,s}(P)^{+1} \subseteq P_t^{+1}(s)$.
Moreover, since Dynamic Consistency holds, we can proceed as in the proof of Theorem 2 to obtain the equality $\varphi_{t,s}(P) = \tilde{\varphi}_{t,s}(P)$ for any $t \in \{0, \ldots, T-1\}$, $s \in S$ and $P \in \mathcal{P}$ such that $\pi_t(s)$ is not P-negligible.

Now we use these facts to show by induction that for any $t \in \{0, \ldots, T\}$ and $s \in S$ such that $P_t(s) \neq \varnothing$

$$\varphi_{t,s}(P) \ \subseteq \ \mathtt{rect}_{t,s}(P) \tag{21}$$

First, if $t = T$, we have $\varphi_{T,s}(P) \subseteq P_T(s)$ which shows (21). If $t = T-1$, we have

$$\varphi_{T-1,s}(P) = \tilde{\varphi}_{T-1,s}(P) = \{ \int_S p(s') \cdot dm(s') : m \in \varphi_{T-1,s}(P)^{+1}, \ p(s') \in \varphi_{T,s'}(P) \ \}$$

But we know that $\varphi_{T-1,s}(P)^{+1} \subseteq P_{T-1}^{+1}(s)$ if $P_{T-1}(s) \neq \varnothing$ and $\varphi_{T,s'}(P) \subseteq P_T(s')$. Therefore, we obtain

$$\varphi_{T-1,s}(P) \ \subseteq \ \{ \int_S p(s') \cdot dm(s') : m \in P_{T-1}(s)^{+1}, \ p(s') \in P_T(s') \ \} \ = \ \mathtt{rect}_{T-1,s}(P) \tag{22}$$

To complete the proof in the case where $t < T$, we proceed as in the case where $t = T-1$.
As for the necessity of the axioms, Reduction follows directly from the fact that $(\varphi_{t,s})_{t \in \{0,\ldots,T\}}^{s \in S}$ is grounded. Last, to show Local Dominance, first note that, since $P_t(s) \neq \varnothing$, we have $\varphi_{t,s}(P) \subseteq \mathtt{rect}_{t,s}(P)$. Thus, $\varphi_{t,s}(P)^{+1} \subseteq \mathtt{rect}_{t,s}(P)^{+1} = P_t(s)^{+1}$. Now, take $f, g \in \mathcal{F}$ that are π_{t+1}-measurable such that $(\{p\}, f) \succsim_{t,s} (\{p\}, g)$ for any $p \in P_t(s)$. Since $\varphi_{t,s}(\{p\}) = \{p\}$ (by $p \in P_t(s)$ and groundedness), $\mathbb{E}_p(u \circ f) \geq \mathbb{E}_p(u \circ g)$ for any $p \in P_t(s)$ and, therefore, for any $p \in \varphi_{t,s}(P)^{+1} \subseteq P_t(s)^{+1}$. But then, since f, g are π_{t+1}-measurable, we have $\mathbb{E}_p(u \circ f) \geq \mathbb{E}_p(u \circ g)$ for any $p \in \varphi_{t,s}(P)$. This finally shows $U_{t,s}(P, f) \geq U_{t,s}(P, f)$. Hence Local Dominance.

[1] Delbaen, F. (2002): "Coherent risk measures on general probability spaces'," *Advances in finance and stochastics*, pp. 1–37.

[2] Ellsberg, D. (1961): "Risk, ambiguity, and the Savage axioms," *Quarterly Journal of Economics*, 75, 643–669.

[3] Epstein, L., and M. LeBreton (1993): "Dynamically consistent beliefs must be Bayesian," *Journal of Economic Theory*, 61(1), 1–22.

[4] Epstein, L., and M. Schneider (2003): "Recursive multiple prior," *Journal of Economic Theory*, 113, 1–31.

[5] Föllmer, H., and I. Penner (2006): "Convex risk measures and the dynamics of their penalty functions," *Statistics and Decisions*, 24(1), 61–96.

[6] Gajdos, T., T. Hayashi, J.-M. Tallon, and J.-C. Vergnaud (2008): "Attitude toward imprecise information," *Journal of Economic Theory*, 140, 23–56.

[7] Gilboa, I., and D. Schmeidler (1989): "Maxmin expected utility with a non-unique prior," *Journal of Mathematical Economics*, 18, 141–153.

[8] Hanany, E., and P. Klibanoff (2007): "Updating preferences with multiple priors," *Theoretical Economics*, 2(3), 261–298.

[9] Hill, B. (2016): "Dynamic Consistency and Ambiguity: A Reappraisal," mimeo.

[10] Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Dynamic variational preferences," *Journal of Economic Theory*, 128(1), 4–44.

[11] Pires, C. (2002): "A Rule For Updating Ambiguous Beliefs," *Theory and Decision*, 53(2), 137–152.

[12] Riedel, F. (2004): "Dynamic Coherent Risk Measures," *Stochastic Processes and Their Applications*, 112(2), 185–200.

[13] Sarin, R., and P. Wakker (1998): "Dynamic Choice and Nonexpected Utility," *Journal of Risk and Uncertainty*, 17(2), 87–119.

[14] Siniscalchi, M. (2011): "Dynamic choice under ambiguity," *Theoretical Economics*, 6(3), 379–421.